DTU Library

The road and its influence on bicycle accidents in Denmark

Janstrup, Kira Hyldekær; Møller, Mette; Pilegaard, Ninette

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

The road and its influence on bicycle accidents in Denmark

Kira H. Janstrup a*, Mette Møller a, Ninette Pilegaard a

[a] *Technical University of Denmark,, Bygningstorvet Building 116, Kgs. Lyngby, 2800, Denmark*

Abstract

The National accident database is often used as basis when designing and prioritizing safety initiatives for cyclists. Due to the very low reporting rate this is not optimal. The purpose of this study is to get a better understanding of factors influencing the occurrence of cyclist accidents with a particular focus on the influence of the condition of the road. The study is based on data on cyclist injuries reported to the hospital and merged with road data, including information on road condition and existence of bicycle lane. The data is analyzed using a Latent Class Clustering approach for pattern recognition. The analysis uncovers patterns of road maintenance and cyclists accidents and reveals 11 clusters. The results identify the road condition as a significant factor for many of the accidents, especially for accidents involving less experienced cyclists. In addition, the analysis confirms that the use of medical records together with road maintenance data leads to new insight of the occurrence of bicycle accidents, which is relevant for the prioritization of preventive efforts.

Keywords: Cyclist accidents; Road maintenance; Medical records; Latent Class Clustering.

* Corresponding author. Tel.: 45 25 65 26
E-mail address: kija@dtu.dk

1. Introduction

Every year, 17,500 cyclist seeks medical care at the hospitals in Denmark following an accident, but several studies have shown that only a small part of these cyclist accidents appears in the National database (Juhra et al, 2012; Janstrup et al., 2016). The National accident database, which consist of all the police reported accidents, is often used as the basis when designing and prioritizing safety measures for cyclists. Due to the heterogeneity in the reporting rate and the very low rate, this is not an optimal solution and may lead to less relevant road safety initiatives for cyclists. Recent studies have shown that the official road crash statistics in Denmark catches about 14% of the severe cyclist accidents and only 6 % of the cyclist accidents with a slight injury (Janstrup et al., 2016). The same study showed that the reporting rate increases with the number of vehicles involved, which means that especially the single cyclist accidents suffer from under-reporting. Beck et al. (2016) have shown, that more than half of the cyclist accidents, which occurred on roads are single accidents. At the same time, it have been shown that these single-cyclist accidents were more likely to occur in the dark, under wet conditions and in rural areas compared to multiple cyclist accidents (Boufous et al., 2013). A study by de Geus et al. (2012) has even concluded that in order to decrease the number of cyclist accidents, measures should be taken to maintain a clean cycling surface and decrease the number of obstacles on bicycle infrastructure. Studies have shown that a high percentage of injured cyclists in single accidents are seriously injured (BoeleVos et al., 2017; Orsi et al., 2017) and that cyclists above 75 years old more often experience a fall from a bicycle than younger cyclists. The number of bicyclists is increasing in Europe and therefore, bicycle safety and knowledge on cycling related road traffic accidents is increasingly important. In order to efficiently improve safety for cyclists a better understanding of cyclist accidents and their contributing factors is needed. Further, high quality data sources with a higher reporting rate is necessary. To achieve this, it is relevant to include additional data such as medical records and road maintenance data. While risk factors such as driving under the influence of alcohol (Orsi et al., 2014), speed (Kim et al., 2007), cycle infrastructure (Teschke et al., 2014; Møller and Hels, 2008; Jensen, 2017) are well documented only limited knowledge on the influence of maintenance level exist. Until now, this has partly been due to a lack of access to relevant data or even a lack of interest in the topic.

Focusing on other transport modes than cyclists, studies have shown that a low road maintenance level may lead to traffic disruption and increase accident risk (Pulugurtha et al., 2013; Corazza et al., 2016). In line with this, survey based studies have shown that road condition and design is an accident factor for the majority of cycling accidents (see NHTSA, 2012). Even naturalistic studies have concluded that poor road surface increases the risk for a cyclist accident tenfold (Dozza and Werneke, 2014). This shows the need for more research on the subject and the relevance of a specific focus on influence of aspects such as road condition on the occurrence of cyclist accidents and the severity of these.

The current study has two main purposes. The first is to get a better understanding of cyclist accidents and contributing factors, with a particular focus on the influence of the road conditions, cycling traffic volume and infrastructure elements. The second aim is to examine if particular road conditions contribute to cycling related injury among subgroups of cyclists. To investigate this a Latent Class Clustering approach is used for pattern recognition. The study is based on data on cyclist injuries reported to the hospital merged with road data including information on road condition and existence of bicycle lane together with the annual daily cycling traffic

The remainder of the paper is organized as follows. Section 2 presents the data and describes the variables used in the analysis. Section 3 presents the method and section 4 presents the result. Last, section 5 offers a discussion and concluding remarks.

2. Data

The data used in this study consists of medical records from road users in Aarhus[†], which have visited the emergency room in the years 2010 to 2015. The data consist of 9,446 observations (road users) of which 4,205 were cyclists. To get more information about the road design and maintenance level we merge the medical records with road maintenance data, collected in the same period, 2010-2015. The road maintenance data includes 9,214 observations with detailed information on 1,567 road sections. Some of the observations of the injured cyclists

[†] Aarhus is the second largest municipality in Denmark with a population of approximately 320.000.

could not be merged with road data, due to missing information, on accident location, and therefore we end up with 3,324 observations in our merged dataset.

The medical records originates from the University hospital of Aarhus and the accident data are reported by nurses at the hospital who gather information on how and why the accident happen by questioning injured road users and medical staff who have been present at the crash scene. The data includes accident characteristics; mode types involved (e.g. car, moped, cyclist), accident type (e.g. single accident, pedestrian accident) and accident location (e.g. intersection, road design). Information on accident circumstances is also listed (e.g. condition of the surface, weather condition). Finally, some road user specific information is listed; age, gender, helmet use and injury severity. The injuries were classified according to severity: no injury or suspected injury (22%), slight injury (59%), severe injury (18%), and fatality (1%).

Road maintenance data hold information on different road characteristics; existence of a bicycle lane and its condition, potholes, rutting, crazing, patching, depression and fretting. These road problems is given on a severity scale based on the size of the damage in the given road section. The road damages were also used to calculate a "damage point" for each road section, where the severity of the road damage defines how much each type of damage will weight in the calculation of the "damage point". For the purpose of the current study, the road condition is defined by the value of the "damage point". The condition of the road is defined as good (new road) for "damage point" between 0 and 1.0, acceptable for damage point between 1.1 and 4.0, poor for damage point above 4.0. Further, information on curbstone where a height above 7cm is defined as too high, shoulder jump (whether or not the shoulder is in line with the road) and at last if the shoulder fall is away from the road or have an angle leading water into the road.

To link the two databases a unique road identification number, the date for the latest road inspection and the date for the occurrence of the accident was used. The linking procedure was made in the software SAS. For all linked observations, a categorization of the annual daily cycling traffic was made. The categorization was based on bicycle census gathered by the municipality of Aarhus. For 51 observations no bicycle census existed. In these cases an expert from the municipality of Aarhus categorized the observations. The categorization resulted in the following groups: 1. 0-500; 2. 501-1500; 3. 1501-3000; 4. 3001-5000; and 5. more than 5001.

3. Method

The data in this study is analyzed using a Latent Class Clustering approach, which is also known as a finite-mixture clustering or a model-based clustering. The approach is used in this study for pattern recognition of the merged cyclist injuries and maintenance data. Latent Class Clustering can be considered as an unsupervised learning approach as the number of clusters and their form is unknown (e.g., Depaire et al., 2008; Magidson and Vermunt, 2002). The main advantages of the Latent Class Clustering over more traditional clustering approaches (e.g., k-means clustering, hierarchical clustering) are: first, its ability to represent overlap across clusters; second, the existence of an underlying statistical model that allows calculating cluster probabilities; and last, the provision of several goodness-of-fit criteria, which makes the decision of the optimal number of clusters much more simple (see, e.g., Depaire et al.,2008; Magidson and Vermunt, 2002).

The Latent Class Clustering technique makes a classification of similar objects into *C* latent classes, by assuming that *N* observations forms a vector which are characterized by another vector of *M* variables $(y_i = y_1,...,y_M)$, and let $(Y_i = Y_{i1},...,Y_{iM})$ be the vector of values of observation *i* for the *M* items. Then, the latent class (clustering) model is given by the following (e.g. Depaire et al., 2008; Kaplan and Prato, 2013):

$$p(Y_i \mid \theta) = \sum_{k=1}^{K} P(C_k) p(Y_i \mid C_k, \theta_k) \tag{1.1}$$

where *K* is the number of clusters, $(k = 1,...,K)$ indicate a cluster, $P(C_k)$ denotes the prevalence of C_k, θ_k is a vector to be estimated, and $p(Y_i \mid C_k, \theta_k)$ denotes the conditional multivariate probability that an observation in class C_k would be characterized by Y_i.

The model formulation is very flexible but in order to reach an estimable model some simplified assumptions is made. Firstly, to meet the nature of the variables in accident data, every variable *i* is assumed to be an ordinal indicator with R_m possible responses $(r_{mi} = 1,\ldots,R_{mi})$. Secondly, since none of the cyclists appears in more than one accident, it is assumed that the observations are uncorrelated. Thirdly, the categorical indicators are independent within a cluster, hence the within-cluster covariance matrix were assumed to be diagonal. Last, all of the categorical indicators were assumed to be endogenous, hence no covariates were employed to predict cluster membership. Under these conditions the Latent Class Clustering model can be formulated by the following (Lanza et al. (2007)):

$$p\left(Y_i \mid \theta\right) = \sum_{k=1}^{K} \pi_k \prod_{m=1}^{M} \prod_{r_m=1}^{R_m} \theta_{mr_m|k}^{I\left(y_{im}=r_m\right)} \tag{1.2}$$

where k is the cluster that observation i is a member of, I is an indicator function that equals 1 if y_{im} equals r_m and 0 otherwise, π_k and θ_{mr_m} are parameters to be estimated. The parameters π_k represent cluster membership probabilities and θ_{mr_m} are indicator response probabilities conditional on the cluster membership. A variation of information criteria (e.g., Akaike information criterion, consistent Akaike information criterion, Bayesian information criterion (BIC), adjusted BIC), could be chosen to decide on the number of clusters. In this study the BIC are used, due to its superiority in terms of consistency and accuracy (see e.g., Nylund et al., (2007)). The Latent Class Clustering technique is performed by using the SAS procedure developed by Lanza et al. (2007).

4. Results

The Latent Class Clustering analysis was performed by using the categorical indicators corresponding to accident and road characteristics. The cyclist injury severity is an outcome of the accident and relates to the cluster characteristics and for that reason it could not be used as an independent variable in the analysis. The variables "helmet use" and "alcohol" is not used either since these two variables are unknown for many of the observations. According to the BIC the Latent Class Clustering yielded 11 clusters. The entropy criterion for this solution was 0.89, which indicates a high certainty in the classification (see, e.g., Depaire et al., 2008). Table 1 describes the characteristics and the prevalence of the clusters while Figure 1 illustrate the latent class characteristics. Table 2 shows the severity, helmet and alcohol distribution in each cluster.

Table 1. Latent class characteristics (percentage of cluster observation)

			C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Variable		Sample	15.3	5.3	13.6	15.5	8.7	11.5	16.5	4.8	2.8	2.0	3.8
Age group	0-8 years	1.2	1.4	2.3	0.4	1.7	1.5	0.0	1.3	2.7	1.1	0.0	0.0
	9-17 years	7.3	3.1	18.7	8.7	5.7	8.0	4.5	10.5	11.5	5.9	0.0	2.4
	18-24 years	27.2	24.6	5.2	24.1	33.1	37.0	30.9	20.1	30.3	20.0	47.4	42.5
	25-39 years	27.5	32.2	18.2	27.1	30.6	24.7	28.5	24.4	21.2	26.4	41.1	27.6
	40-49 years	12.9	13.1	20.2	11.2	12.1	15.5	12.5	13.2	14.3	14.6	1.4	7.9
	50-65 years	18.1	19.4	25.1	19.4	14.6	10.5	16.6	22.4	16.7	26.4	10.1	15.7
	66 + years	5.9	6.2	10.3	9.0	2.2	2.9	7.0	7.9	3.4	5.6	0.0	3.9
Gender	Male	49.7	51.7	62.4	47.0	51.3	42.6	44.8	52.3	54.6	48.0	38.0	48.8
Accident situation	Single	51.9	73.1	68.2	54.9	1.3	57.7	68.8	62.8	53.4	49.0	35.0	37.0
	Same road without turn	14.6	19.6	16.8	15.0	5.0	10.9	6.0	25.3	11.2	19.4	13.4	17.3
	Same road with turn	17.8	2.8	5.3	14.3	55.7	18.8	11.5	1.3	16.5	22.1	26.6	35.4
	Intersection	12.9	2.4	8.8	14.6	38.0	10.0	10.5	3.9	14.4	5.6	16.1	6.3
	Pedestrian	2.9	2.1	0.9	1.1	0.0	2.7	3.2	6.7	4.5	3.9	8.9	3.9
Season	Winter	19.3	7.8	17.5	19.2	12.0	20.1	69.4	3.3	16.6	14.4	26.5	16.5
	Spring	23.7	24.5	24.1	24.3	26.5	25.9	4.8	31.0	19.9	32.4	22.1	27.6
	Summer	27.8	37.9	39.0	28.8	25.9	24.0	0.3	38.6	28.9	30.2	16.1	23.6
	Fall	29.2	29.8	19.4	27.6	35.6	29.9	25.5	27.1	34.6	23.0	35.3	32.3
Road condition	Dry	50.3	50.5	48.8	44.9	66.6	40.1	10.8	70.9	51.0	56.5	48.5	53.5
	Slippery	25.3	14.6	20.7	25.1	19.2	28.6	74.5	7.5	25.4	17.9	24.8	26.0
	Unknown	24.4	34.9	30.4	30.0	14.2	31.3	14.7	21.6	23.7	25.6	26.7	20.5
Light	Daylight	66.7	67.9	72.5	62.8	77.9	64.5	28.0	83.6	62.0	81.3	57.3	71.7
	Dark	32.3	31.0	24.9	35.7	21.9	34.1	72.0	15.9	36.8	15.4	39.7	27.6
	Unknown	1.0	1.1	2.6	1.5	0.1	1.4	0.0	0.5	1.3	3.3	3.0	0.8
Road design	Bicycle lane	18.9	22.2	10.9	9.7	0.3	13.5	23.7	37.3	19.1	25.0	32.1	32.3
	Straight	36.5	54.8	54.6	50.1	0.0	38.6	35.2	43.7	34.0	34.1	23.7	16.5
	Intersection	36.6	15.2	16.6	32.0	89.5	36.0	35.8	13.3	37.5	36.1	44.2	50.4

	Curve	1.8	2.1	7.4	2.1	0.1	2.4	1.7	1.4	2.4	1.1	0.0	0.0
	Roundabout	1.5	1.3	0.7	1.6	5.4	0.6	1.2	0.2	0.0	0.0	0.0	0.0
	Other	4.8	4.4	9.9	4.6	4.6	8.9	2.5	4.1	7.0	3.7	0.0	0.8
AADT bicyclist	0-500	16.0	4.3	86.2	14.4	16.1	13.3	15.1	18.0	6.4	0.0	6.7	0.0
	501-1500	20.4	19.6	0.0	12.0	23.0	24.1	19.7	22.6	27.3	99.4	1.4	0.0
	1501-3000	16.6	16.9	13.0	7.4	18.9	27.0	19.8	17.0	24.3	0.0	40.0	0.0
	3001-5000	17.2	10.0	0.0	12.5	11.1	16.5	11.3	17.8	34.5	0.5	51.9	99.9
	5001-	29.7	49.2	0.8	53.7	30.9	19.2	34.1	24.5	7.5	0.1	0.0	0.0
Bicycle lane condition	Good	21.4	46.9	13.9	2.4	20.3	1.4	9.1	1.2	38.2	67.7	51.8	99.9
	Acceptable	10.4	16.7	4.4	11.2	10.2	18.9	6.8	2.5	20.0	24.4	0.0	0.0
	Bad	3.3	0.5	1.9	7.9	0.3	12.3	0.0	0.0	0.2	0.0	47.9	0.0
	No record	64.9	35.9	79.8	78.5	69.3	67.4	84.1	96.3	41.7	7.9	0.3	0.0
Road Surface	Good	59.1	84.0	54.0	73.6	79.0	18.4	67.2	66.8	13.8	0.1	0.1	0.0
	Acceptable	37.1	15.9	35.5	26.4	20.5	62.0	29.8	31.7	68.3	99.9	99.9	100
	Bad	3.8	0.1	10.5	0.0	0.4	19.7	3.0	1.4	17.9	0.0	0.0	0.0
Curbstone	High	6.7	0.5	1.8	4.0	0.0	4.7	0.4	0.0	0.0	58.1	6.7	99.9
Drains	Damaged	10.0	0.0	12.1	14.9	0.1	71.4	0.2	0.0	0.0	0.0	51.8	0.0
Peeling	Yes	10.7	0.0	19.8	8.5	0.0	78.0	0.0	0.0	14.7	37.9	0.0	0.0
Crazing	Yes	25.0	5.0	83.7	23.6	4.1	97.6	2.4	0.7	85.4	33.7	99.9	0.0
Patching	Yes	50.5	33.2	89.5	84.1	22.5	99.6	30.7	19.8	92.5	26.5	58.7	100
Depression	Yes	32.7	22.2	44.7	100	8.7	100	3.3	0.0	14.2	5.2	99.9	0.0
Fretting	Yes	57.7	83.8	58.2	98.9	37.8	98.8	24.1	8.4	100	100	99.9	0.0
Potholes	Yes	11.8	0.0	35.9	4.7	1.6	76.9	0.7	0.0	47.2	0.0	0.0	0.0
Cross fall	Yes	21.9	0.0	3.2	83.9	0.0	94.8	0.0	0.0	0.0	0.0	99.9	0.0
Shoulder fall	Away from road	1.4	0.0	25.1	0.0	0.3	0.0	0.0	0.2	0.0	0.0	0.0	0.0
Shoulder	Jump	3.8	0.0	55.8	2.7	0.0	2.8	0.0	0.9	0.0	1.6	0.0	0.0

Figure 1. Illustrative latent class characteristics

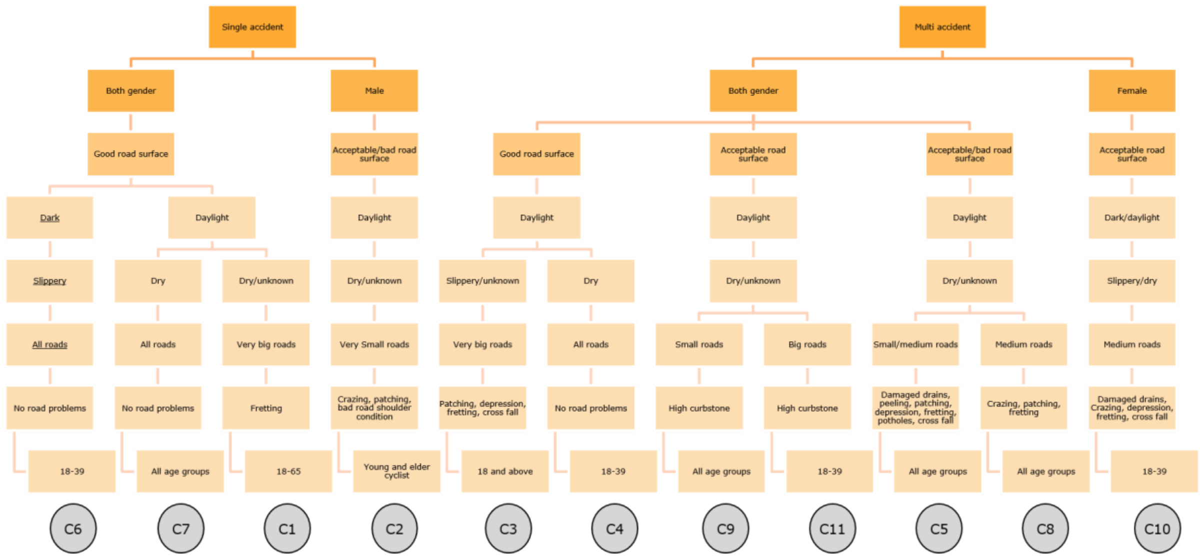

Table 2. Severity, helmet and alcohol characteristics (percentage of cluster observation)

			C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Variable		Sample	15.3	5.3	13.6	15.5	8.7	11.5	16.5	4.8	2.8	2.0	3.8
Severity	Fatal	0.3	0.2	0.0	0.2	1.1	0.0	0.0	0.0	0.0	1.0	0.0	0.0
	Severe	18.5	19.0	23.9	20.4	16.4	21.1	15.3	19.5	13.7	22.7	13.2	15.0
	Slight	59.1	58.0	63.1	58.2	53.1	55.3	63.7	62.1	68.0	54.6	63.2	57.5
	No injury	22.2	22.7	13.1	21.2	29.4	23.6	21.0	18.4	18.3	21.6	23.5	27.6
Helmet	Used	37.2	31.0	56.8	36.1	40.7	29.6	40.7	37.5	35.3	41.2	29.4	33.1
	Not used	52.6	55.7	31.3	53.3	52.9	54.2	52.5	52.1	55.6	45.4	66.2	58.3
	Unknown	10.2	13.3	11.9	10.6	6.4	16.2	6.8	10.4	9.2	13.4	4.4	8.7
Alcohol	Yes	6.2	9.8	2.3	7.7	0.6	7.4	7.9	6.8	7.8	1.0	7.4	5.5
	No	63.7	55.9	59.7	61.3	75.7	62.0	64.5	66.1	54.9	62.9	67.6	59.8
	Unknown	30.1	34.3	38.1	31.0	23.7	30.6	27.6	27.1	37.3	36.1	25.0	34.6

The 11 clusters will be described and characterized in the following sections.

4.1. C1: Single accidents on big and mainly roads in good condition

Cluster C1 is one of the largest clusters consisting of 15.5% of all the observations. The age distribution in this cluster is quite broad and covers cyclists in the age 18 to 65 years old and both genders. 73.1% of all the accidents in this cluster are single accidents which occurred during the cycling seasons (e.g. summer, spring, fall). The road condition at the time of the accident was dry (50.5%) or unknown (34.9%) and it happened in daylight (67.9%). The accidents occurred mainly on roads with an annual daily cycling traffic volume above 5001 (49.2%), which is a very high amount of cyclists. Other than occurring on large roads, the roads were also characterized by a good surface and a bicycle lane in good condition. The only surface problem on the roads was fretting (83.8%). This cluster had the highest percentage of cyclist who were tested positive for alcohol at time of the accident (9.8%).

4.2. C2: Young and old cyclists on roads with shoulder jump and fall

Cluster C2 includes 5.3% of the observations and is composed mainly of male cyclists in the age 0 to 17 years old or above 66 years old. The most frequent type of accident is a single accident (68.2%) occurred in daylight (72.5%) in the summer season (39.0%). The accidents had happened on small roads with an annual daily cycling traffic below 500 (86.2%), which are very few cyclists. On these roads there no records of the bicycle lane. The condition of the road was mixed but with crazing (83.7%), patching (89.5%) and some potholes (35.9%). Furthermore, the accidents in this cluster had occurred on roads whit a high shoulder edge (55.5%) and a wrong shoulder fall which can give problems with water and dirt in the sides of the roads (where the cyclists drives). This cluster also includes the highest percentage of cyclists who use a helmet (56.8%) which is a lot more than the sample percentage (37.2%), but it still include a big amount of cyclists with a severe injury (23.9%).

4.3. C3: Accidents on big and poor roads

Most of the cyclists in this cluster, C3, is older than 18 years old and represents both female (53.0%) and male (47.0%). A large part of the accidents is single accidents (54.9%), but also some intersection accidents (14.9%). The accidents has mainly occurred in daylight (62.8%) on a straight road (50.1%) or in an intersection (32.0%). The roads for the accidents are big with an annual daily cycling traffic above 5001. There is no records of a bicycle lane (78.5%) on these roads and the condition of the surface is mainly good (73.6%). Even though the condition of the surface are mainly good there are records of road damage as patching (84.1%), depression (100.0%) and fretting (98.9%). On the roads where the accidents had occurred there are also registration of cross fall (83.9%) with a fall leading water and dirt out in the side of the roads, where the bicyclist drives.

4.4. C4: Accidents with a turn on very good roads

The cluster, C4, is one of the biggest cluster and include 15.5 % of all cyclist injuries. The cyclist age is mainly between 18 and 39 years old (63.7%). A large part of these accidents has happened in an intersection (89.5%) where the other part has taking a turn (55.7%) and hit the injured cyclist. The accidents has mainly taking place in the fall season (35.6%) and in daylight (77.9%) on dry surface (66.6%). Many of the accidents has taken place on a road without any bicycle lane registration (69.3%). The roads surface condition is very good and no damage had been recorded. This cluster can also be described as the most severe cluster since most fatalities (1.1) were included in this cluster. Furthermore, the lowest percentage of drivers who have been drinking was found in this cluster, namely 0.6%.

4.5. C5: Single accidents with females on very poor roads

8.7% of all the cyclist injuries are included in cluster C5. All age groups are represented and the cyclists are mainly females (57.4%). The accident type are mainly single (57.7%) and had happen on a straight road (38.6%) or in an intersection (36.0%). On those places where a bicycle lane exist, the condition of the bicycle lane was acceptable (18.9%) or bad (12.3%). The highest amount of cyclists for whom the accident had happen on a road with an unacceptable road surface was included in this cluster (19.7%). For that reason, all kind of road damages appear on these roads (e.g. peeling, crazing, patching, depression, fretting, potholes). Furthermore, there was records on damaged drains (71.4%) and wrong cross fall (94.8%) on these roads.

4.6. C6: Single accidents with females on slippery roads

This cluster includes cyclist above 18 years old and mainly females (55.2%). The injured cyclists have typically been involved in a single accident (68.8%) on a slippery road (74.5%) in some dark hours (72.0%). The accident have happen on all types of road designs and sizes. There are rarely records on the bicycle lanes condition (84.1%) while the condition of the road surface typically have been registered as good (67.2%).

4.7. C7: Accidents on roads with a good surface and with a bicycle lane

This cluster holds the largest amount of injured cyclists namely 16.5%, which includes all age groups and both gender. The accident situation is mainly single (62.8%) or on same road with same direction without turn (25.3%). Almost none of the accidents have happen in the winter season (3.3%) and they have occurred on dry surface (70.9%) in daylight (83.6%). Even though a large amount of the accidents had happen on a bicycle lane (37.3%) there is no records of the condition of the bicycle lane (96.3%). The condition of the roads for where the accident have happen is mainly good (66.8%) and no road damage has been registered.

4.8. C8: Cyclists on roads with an unacceptable road condition

This cluster, C8, is one of the smallest and include 4.8% of all the cyclist injuries. Cyclists in all ages are represented in this cluster where there is a bit more males (54.6%). The cyclists in this cluster have different accidents types where single accident represents the highest part (53.4%). The accident typically happened in daylight (62.0%) on a road with a mixed condition. The road design has been straight (34.0%) or in an intersection (37.5%). The annual daily cycling traffic is between 500 and 5000 hence, the roads for where the accident happened are mixed between small and large roads. The condition of the roads surface is acceptable (68.3%) and unacceptable (17.9%) and many road specific problems have been registered (e.g. crazing (85.4%), patching (92.5%), fretting (100.0%) and potholes (47.2%) at the location where the accident happened. This cluster also includes the highest percentage of cyclists with a slight injury (68.0%).

4.9. C9: Elder cyclists on roads with a high curbstone

This cluster contains the largest amount of cyclists in the age group 50 to 65 years old of both gender. The accident type are single (49.0%), on a same road with a turn (22.1%) and on same road without a turn (19.4%). The accidents happened on a dry road (56.5%) in daylight (81.3%). Again, the road design is a mixture of bicycle lane (25.0%), straight (34.1%) and intersection (37.5%) while the annual daily cycling traffic is low, between 500 and 1500 (99.4%). There exists records of the condition of the bicycle lane and this condition is mainly good (67.7%). The roads also have an acceptable maintenance level and the only registered road problem is fretting (100.0%). Furthermore, problems with a high curbstone are registered for many (58.1%) of the roads for where the accident had occurred.

4.10. C10: Younger females on roads with cross fall

C10 is the smallest cluster and include only 2.0% of the observations. The cyclists in this cluster are mainly females (62.0%) and in the age group 18 to 39 years old (88.5%). All types of accidents are represented here and the accidents have mainly occurred in daylight (57.3%). A high amount of the accidents happened in an intersection (44.2%) and the annual daily cycling traffic are between 1500 and 5000. The condition of the bicycle lane is available for the places where the accidents has occurred but this condition is recorded as both bad (47.9%) and good (51.8%). The condition of the road surface is acceptable (99.9%) and there was a lot of road problems (e.g. crazing, depression and fretting) registered at the time for the accident. Further, for these roads also damaged drains (51.8%) and cross fall (99.9%) were found. This cluster include the lowest percentage of cyclist who use a helmet (29.4%).

4.11. C11: Young cyclists on roads with a high curbstone

The last cluster, C11, found with the Latent Class Clustering technique includes cyclists in the age 18 24 years old (42.5%). The main accident type for the cyclists in this cluster is single accident (37.0%) and same road with turn (35.4%). The road design is bicycle lane (32.3%) or intersection (50.4%) with a high annual daily cycling traffic 3001-5000 (99.9%). The surface condition for the bicycle lane was good, while the road was found to have an acceptable maintenance level. The only road problems found on these roads was patching (100.0%) and that the height of the curbstone (99.9%) was too high (above 7 cm). The highest percentage (27.6%) of cyclists with no or suspected injuries are included in this cluster.

5. Conclusion and discussion

Results uncover patterns of road maintenance and cyclist accidents reported to the emergency room in Aarhus and investigate prevalence and severity of accidents on roads with certain surface conditions. The results shows that the merge of accident data and road data gives important knowledge on the accidents and can give a better understanding of what happens in a cyclist accident. The clustering technique provides a holistic and multidimensional overview of the cyclist accident pattern and gives knowledge for prioritize safety issues for road maintenance and construction projects. The results revealed 11 "cyclist" clusters where 3 of the clusters (C1 "Single accidents on big and mainly good roads", C4 "Accidents with a turn on very good roads" and C7 "Accidents on roads with a good surface and with a bicycle lane") mainly included cyclist which had an accident on a road without any road problems.

The road problems seemed to be an issue for 8 of the identified clusters, which in total include 53% of all the observations. The analyses show that conditions as high curbstone, shoulder jump and wrong shoulder fall was a problem among young (less than 18 years old) and elder (more than 55 years old) cyclists' accidents pattern (C2 "Young and old cyclists on roads with shoulder jump and fall", C9 "Elder cyclists on roads with a high curbstone" and C11 "Young cyclists on roads with a high curbstone"). This could indicate that particularly unsafe and less experienced cyclist indeed have problems when driving in streets with a bad infrastructure. These cyclists were generally not under influence of alcohol but whether or not they had an otherwise more risky behavior than other cyclists cannot be answered by this study.

The analyses also show that female cyclist are exposed on roads with a slippery surface (C6 "Single accidents with females on slippery roads") and with many potholes (C5 "Single accidents with females on very poor roads") and cross fall (C10 "Younger females on roads with cross fall") leading water and dirt out in the side of the roads. A large part of the accidents in cluster C5 and C6 are single accidents and especial this type of accidents are heavily under-reported in the official registers.

The last two cluster identified in the analysis C3 "Accidents on big and poor roads" and C8 "Cyclists on roads with an unacceptable road condition" have a very bad maintenance level, where problems as crazing, patching and fretting were identified. These two clusters include no less than 18.4% of all the cyclist injuries and includes cyclists in all ages and both gender. A better maintenance level could potentially prevent some (or even all) of these accidents, but to examine whether or not this is the case, it is necessary to know more about the contributing accident factors.

This study shows, that the application of emergency room data and other data sources is an important element to get more knowledge on what causes single accidents. The analyses illustrates the importance of merging data sources to get more knowledge on different aspects of road safety, in this case the road. The results of this study conclude that roads maintenance level and infrastructure have an effect on the cyclist accident pattern and that more than 50% of the cyclist injuries have happened on roads with poor maintenance level. It is therefore of highly relevance to continue the research to explore the interaction between the road, the vehicle and the behavior of consequence for cycling accidents.

Acknowledgements

The financial support of the Danish Asphalt Pavement Association and the Danish Safe Roads Association is gratefully acknowledged. The authors would also like to thank the Municipality of Aarhus for expert knowledge and data providing.

6. References

Beck, B., Stevenson, M., Newstead, S., Cameron, P., Judson, R., Edwards, E.R., Bucknill, A., Johnson, M., Gabbe B., 2016. Bicycling crash characteristics: An in-depth crash investigation study. Accident Analysis and Prevention 96, 219–227.

BoeleVos, M.J., Van Duijvenvoorde, K., Doumen, M.J.A., Duivenvoorden, C.W.A.E. Louwerse, W.J.R., Davidse, R.J., 2017. Crashes involving cyclists aged 50 and over in the Netherlands: An in-depth study. Accident Analysis and Prevention, 105, 4–10.

Boufous, S., de Rome, L., Senserrick, T., Ivers, R. Q., 2013. Single- versus multi-vehicle bicycle road crashes in Victoria, Australia. Inj Prev, 19, 358-362.

Corazza, M.V., Mascio, P.D., Moretti, L., 2016. Managing sidewalk pavement maintenance: a case study to increase pedestrian safety. Journal of traffic and transportation engineering, 3, 203-214.

de Geus, B., Vandenbulcke, G., Panis, L. I., Thomas, I., Degraeuwe, B., Cumps, E., Aertsens, J., Torfs, R., Meeusen, R., 2017. A prospective cohort study on minor accidents involving commuter cyclists in Belgium. Accident Analysis and Prevention, 45, 683– 693.

Depaire, B., Wets, G. and Vanhoof, K., 2008. Traffic Accident Segmentation by Means of Latent Class Clustering. Accident Analysis and Prevention; 40; 1257-1266.

Dozza, M. and Werneke, J., 2014. Introducing naturalistic cycling data: What factors influence bicyclists' safety in the real world? Transportation Research Part F 24, 83–91.

Janstrup, K. H., Kaplan, S., Hels, T., Lauritsen, J., Prato, C. G., 2016. Understanding Traffic Crash Under-Reporting: Linking Police and Medical Records to Individual and Crash Characteristics. Traffic Injury Prevention; 17; 6; 580-584.

Jensen, S. U., 2017. Safe roundabouts for cyclists. Accident Analysis and Prevention 105, 30–37.

Juhra, C., Wieskötter, B., Chu, K., Trost, L., Weiss, U., Messerschmidt, M., Malczyk, A., Heckwolf, M., Raschke, M., 2012. Bicycle accidents – Do we only see the tip of the iceberg? A prospective multi-centre study in a large German city combining medical and police data. Injury, Int. J. Care Injured, 43, 2026–2034.

Kaplan, S. and Prato C.G., 2013. Cyclist-Motorist Crash Patterns in Denmark: A Latent Class Clustering Approach. Traffic Injury Prevention, 14; 725-733.

Kim, J-K., Kim, S., Ulfarsson, G.F., Porrello, L. A., 2007. Bicyclist injury severities in bicycle–motor vehicle accidents. Accident Analysis and Prevention 39, 238–251.

Lanza, S.T., Collins, L.M., Lemmon, D.R. and Schafer, J.L., 2007. PROC LCA: A SAS Procedure for Latent Class Analysis. Structural Equation Modeling, 14, 671-694.

Magidson, J. and Vermunt, J. K., 2002. Latent class models for clustering: A comparison with K-means. Canadian Journal of Marketing Research, 20.1, 36-43.

Møller, M. and Hels, T., 2008. Cyclists' perception of risk in roundabouts. Accident Analysis and Prevention 40, 1055–1062.

N.H.T.S.A. (NHTSA), 2012. National Survey on Bicyclist and Pedestrian Attitudes and Behaviors. National Highway Traffic Administration, 2012.

Nylund, K. L., Asparouhov, T., Muthén, B. O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equation Modeling, 14, 535-569.

Orsi, C. Ferraro, O.E., Montomoli, C., Otte, D., Morandi, A., 2014. Alcohol consumption, helmet use and head trauma in cycling collisions in Germany. Accident Analysis and Prevention, 65, 97– 104.

Orsi, C., Montomoli, C., Otte, D., Morandi, A., 2017. Road accidents involving bicycles: configurations and injuries, International Journal of Injury Control and Safety Promotion, DOI: 10.1080/17457300.2016.1278239

Pulugurtha, S.S., Ogunro, V., Pando, M. A., Patel, K. J., Bonsu, A., 2013. Preliminary results towards developing thresholds for pavement condition maintenance: safety perspective. Procedia – social and behavioral science, 104, 302-311.

Teschke, K., Frendo, T., Shen, H., Harris, M. A., Reynolds, C.C., Cripton, P. A., Brubacher, J., Cusimano, M. D., Friedman, S. M., Hunte, G., Monro, M., Vernich, L., Babul, S., Chipman, M., Winters, M., 2014. Bicycling crash circumstances vary by route type: a cross-sectional analysis. BMC Public Health, 14, 1205.